

Proudly Operated by Battelle Since 1965

Enabling Re-executable Workflows with Near-real-time Visualization, Provenance Capture and Advanced Querying for Mass Spectrometry Data

MATHEW THOMAS

Pacific Northwest National Laboratory
2016 New York Scientific Data Summit (NYSDS)





Topics

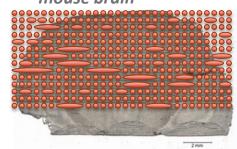
- Nano-DESI Mass Spectrometry experimental setup
 - How is data acquired?
 - Data structure and its complexities
- MSI QuickView Software Toolkit
 - Capabilities
 - Web App
- Workflow components that enable ...
 - Re-execution and reproducibility of experiments
 - ProvEn
 - Collaborative querying of experimental data/information
 - ELK Stack (Elasticsearch, Logstash, Kibana)

Near Real Time Analysis of Mass Spectrometric Data during Acquisition



Proudly Operated by Battelle Since 1965

The optical image of a mouse brain

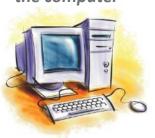


Line 3 Line 4 Line 5 Line 6

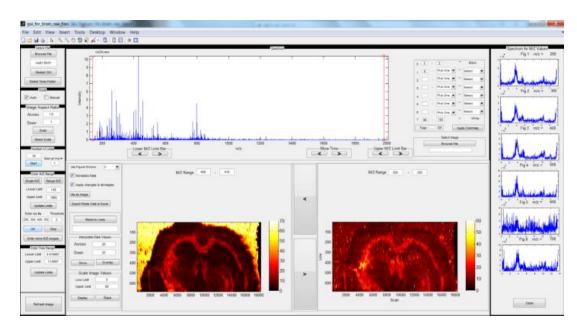
Line 7 Line 8 Line 9 Line 10 Line 11 Line 12 Line 13 Acquire data for each Line (marked) of the optical image using the instrument



Raw File saved onto the computer



.raw

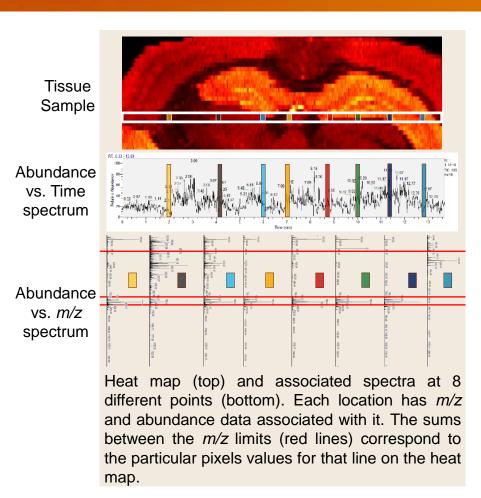


Data for all Lines is
displayed in MSI
QuickView Real-Time
Visualization Tool



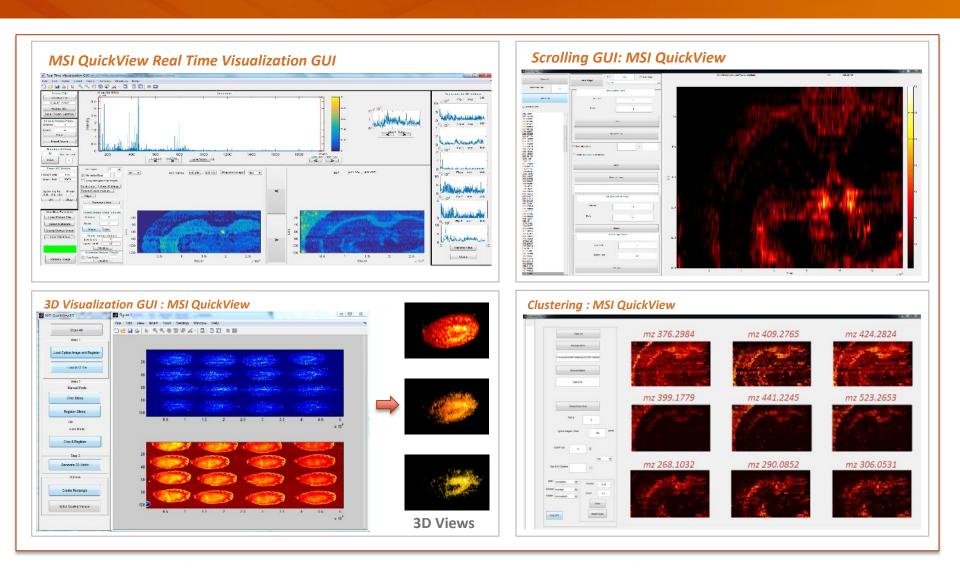
Data Structure

- Mass Spectrometric Imaging (MSI) generates spatial maps that can be used to visualize the location and quantity of molecules.
- Each point in space can have several thousand mass-to-charge (m/z) values.



MSI QuickView Capabilities



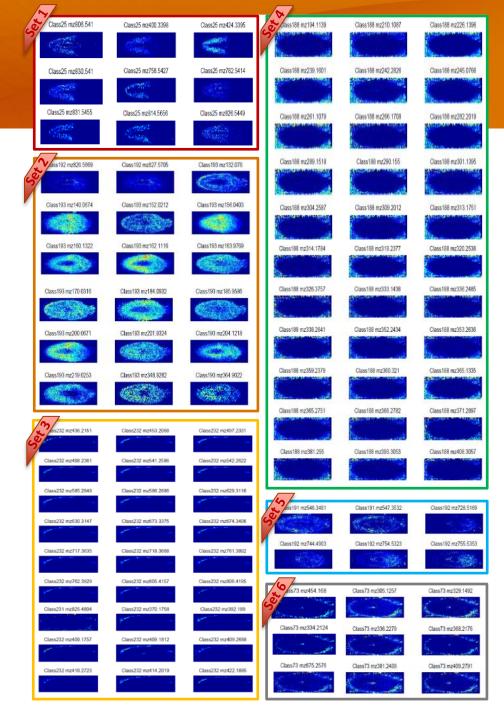


Lanekoff .I.T, Thomas M., and Laskin J.. 2014. "Shotgun Approach for Quantitative Imaging of Phospholipids Using Nanospray Desorption Electrospray Ionization Mass Spectrometry ." Analytical Chemistry 86(3):1872-1880.

Clustering Tool

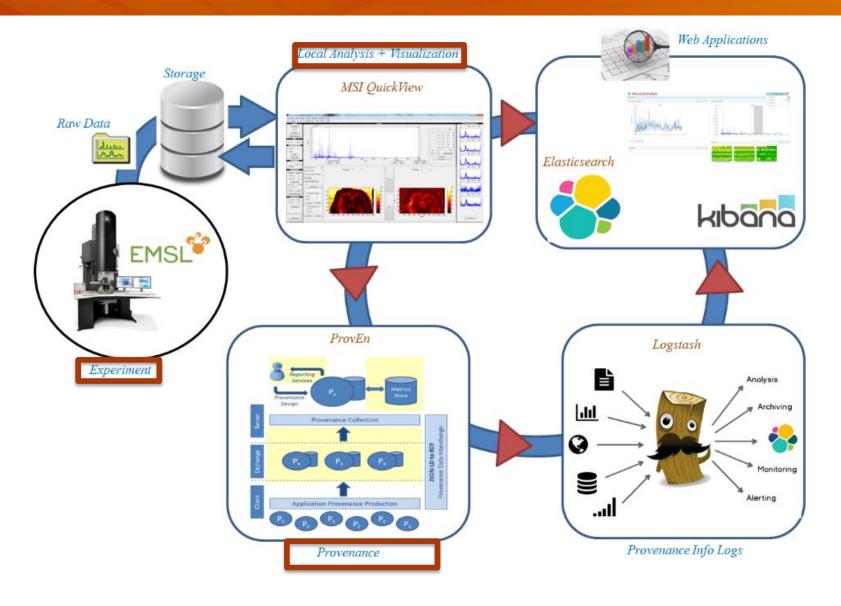
Classifying High Resolution Mass-Spectrometry Data

- ☐ Classification is one of the fundamental methodologies for analyzing mass spectral data.
- ☐ Major goals of classifying massspectrometry data include:
 - Automatically group compounds based on their mass spectra.
 - To determine correlation between properties of compounds and their mass spectra.
- ☐ 6 different classes from a single dataset are shown as an example



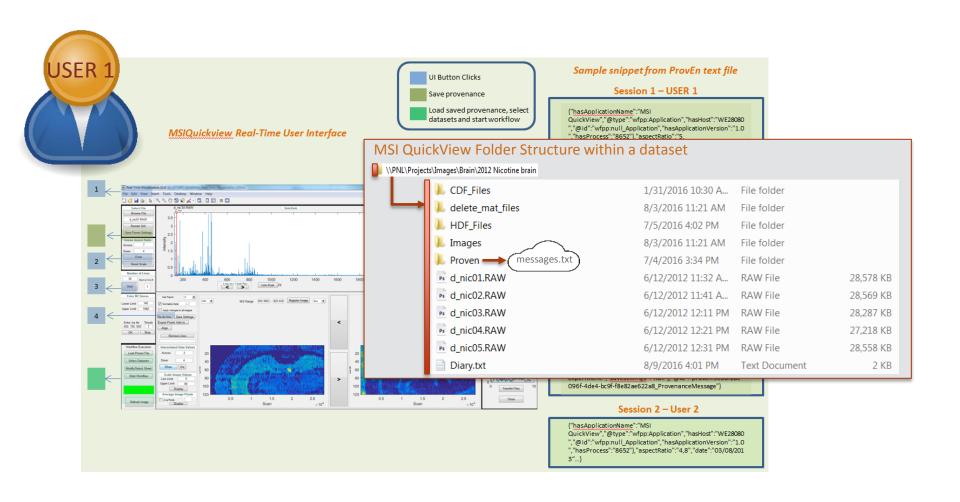
The Workflow





MSI QuickView – Provenance Capture





MSI QuickView – Provenance Capture



Proudly Operated by **Battelle** Since 1965



What does the provenance message look like?

d":"msiquickview:scaleImageValues"},"proven":"http://provenance.pnnl.gov/ns/proven#","applyChangesToAllImages";("@type":"xsd:string","@id":"msiquickview:applyChangesToAllImages"},"hasApplication";("@ ype":"xsd:string","@id":"msiquickview:date"},"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:interpolatedDataValues"},"saveImage":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"msiquickview:date"),"interpolatedDataValues":("@type":"xsd:string","@id":"xsd:string","date:" "normalizeData";{"@type":"xsd:string","@id":"msiquickview:normalizeData"},"redoImage";("@type":"xsd:string","@id":"msiquickview:redoImage"},"redoImageExcelfileName";("@type":"xsd:string","@id":"msiquickview:redoImageTh ckview:redoImageExcelfileName"},"dpiVal":{"@type":"xsd:string","@id":"msiquickview:dpiVal"},"removeLines":{ "@type":"xsd:string","@id":"msiquickview:removeLines"},"mzPlotValsThresh":{"@type":"xsd:string","@id":"msiquickview:mzPlotValsThresh"},"folderLocation":{"@type":"xsd:string","@id":"msiquickview:folderLoc ation"},"updateTime":{"@type":"xsd:long","@id":"wfpp:updateTime"},"numRawFiles":{"@type":"xsd:long","eid":"msiquickview:numRawFiles"},"redoImageExcelSheetName":{"@type":"xsd:string","@id":"msiqui ckview:redoImageExcelSheetName"}, "uniqueID":{"@type":"xsd:string", "@id":"msiquickview:uniqueID"}, "mzRange":{"@type":"xsd:string", "@id":"msiquickview:mzRange"}, "redoImagePDFno":{"@type":"xsd:string" ,"@id":"msiquickview:redoImagePDFno"},"alignImage":{"@type":"xsd:string","@id":"msiquickview:alignImage"},"scientistName":{"@type":"xsd:string","@id":"msiquickview:scientistName"},"redoImageExceImzRo ws":{"@type":"xsd:string","@id":"msiquickview:redoImageExceImzRows"},"includeAxisImageSave";("@type":"xsd:string","@id":"msiquickview:includeAxisImageSave"},"rawStartNo":("@type":"xsd:string","@id":" msiquickview:rawStartNo"},"datasetName":{"@type":"xsd:string","@id":"msiquickview:datasetName"},"exportPixelsValsToExcel":{"@type":"xsd:string","@id":"msiquickview:exportPixelsValsToExcel"},"msiquickview:exportPixelsValsToExcel"} w":"http://msiquickview.pnnl.gov#","hasProvenance";{"@type":"@id","@id":"proven:hasProvenance"},"xsd":"http://www.w3.org/2001/XMLSchema#","hasHost";("@type":"xsd:string","@id":"wfpp:hasHost"},"im ageListToSave":{"@type":"xsd:string","@id":"msiquickview:imageListToSave"},"notes":{"@type":"xsd:string","@id":"msiquickview:notes"},"saveSettings":{"@type":"xsd:string","@id":"msiquickview:saveSettings"}," hasProcess":{"@type":"xsd:integer","@id":"wfpp:hasProcess"}},"@type":"proven:ProvenanceMessage","hasProvenance":{"mzPlotVals":"null","scaleImageValues":"null, null","applyChangesToAllImages":"1","hasApplication":{"hasApplicationName":"MSI

QuickView", "@type": "wfpp:Application", "hasHost": "WE28080", "@id": "wfpp:null Application", "hasApplication": 1.0", "hasProcess": "8652", "aspectRatio": "5,

10","date":"06/12/2012","interpolatedDataValues":"2, 2","saveImage":"null","normalizeData":"1, ,

","redoImage":"C://Users//thom991//Desktop//MSI_testData//","redoImageExcelfileName":"Class_List.xls","dpiVal":"null","colorMap":"16","removeLines":"null","mzPlotValsThresh":"null","folderLocation":"C://U sers//thom991//Desktop//MSI testData//testset1//","updateTime":"1467671381","numRawFiles":"5","@type":"msiquickview:Start","redoImageExcelSheetName":"2","mzRange":"140, 1980", "uniqueID": "enter scaling values Callback(hObject, eventdata, handles), close multiple mz panel Callback(hObject, eventdata, handles), popupmenu for colormap3d Callback(hObject, eventdata, handles),

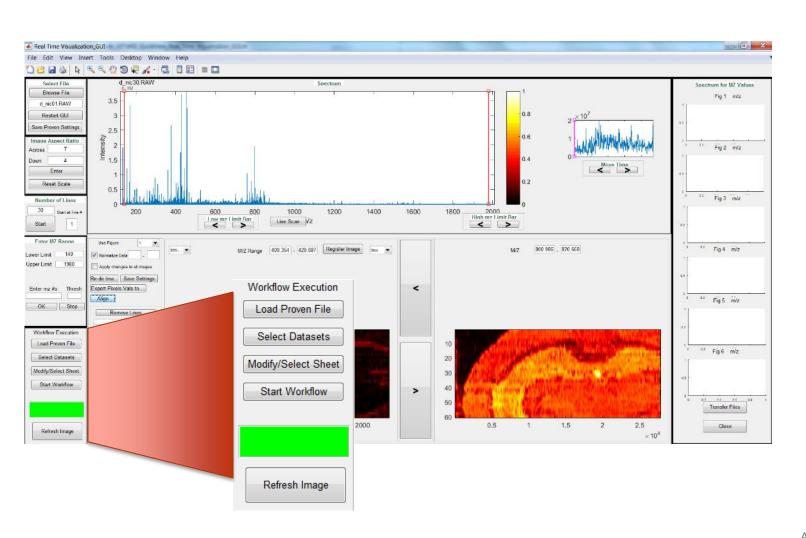
start Callback(hObject, eventdata, handles), apply manipulations to all images Callback(hObject, eventdata, handles), redo for new ranges Callback(hObject, eventdata, handles), submit prov Callback(hObject,eventdata,handles)","redoImagePDFno":"1","alignImage":"null","scientistName":"Julia Laskin","redoImageExcelmzRows":"1,

6","includeAxisImageSave":"null","rawStartNo":"null","datasetName":"Mouse Brain

Nicotine", "exportPixelsValsToExcel": "null", "@id": "msiquickview:da39a3ee5e6b4b0d3255bfef95601890afd80709_Start", "imageListToSave": "null", "notes": "lmaging mass spectrometry offers simultaneous spatially resolved detection of drugs, drug metabolites, and endogenous substances in a single experiment. This is important when evaluating effects of a drug on a complex organ system such as the brain, where there is a need to understand how regional drug distribution impacts function. Nanospray desorption electrospray ionization, nano-DESI, is a new ambient technique that enables spatially resolved analysis of a variety of samples without special sample pretreatment. This study introduces an experimental approach for accurate spatial mapping of drugs and metabolites in tissue sections by nano-DESI imaging. In this approach, an isotopically labeled standard is added to the nano-DESI solvent to compensate for matrix effects and ion suppression. The analyte image is obtained by normalizing the analyte signal to the signal of the standard in each pixel. We demonstrate that the presence of internal standard enables online quantification of analyte molecules extracted from tissue sections. Ion images are subsequently mapped to the anatomical brain regions in the analyzed section by use of an atlas mesh deformed to match the optical image of the section. Atlas-based registration accounts for the physical variability between animals, which is important for data interpretation. The new approach was used for mapping the distribution of nicotine in rat brain tissue sections following in vivo drug administration. We demonstrate the utility of nano-DESI imaging for sensitive detection of the drug in tissue sections with subfemtomole sensitivity in each pixel of a 27 ?m × 150 ?m area. Such sensitivity is necessary for spatially resolved detection of low-abundance molecules in complex matrices.", "saveSettings": "null"}, "@id": "proven:9e3df2de-096f-4de4-bc9f-f8e82ae622a8 ProvenanceMessage"}

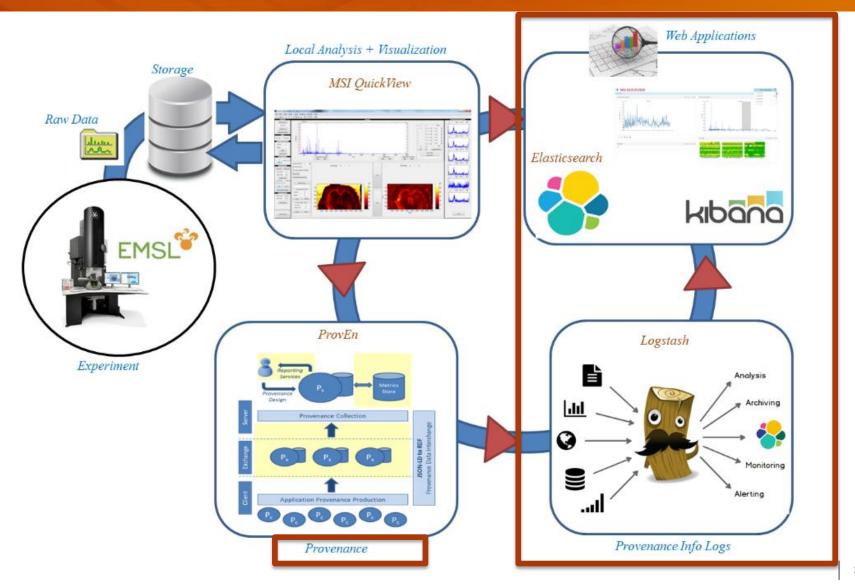
Re-executing Workflows





The Workflow





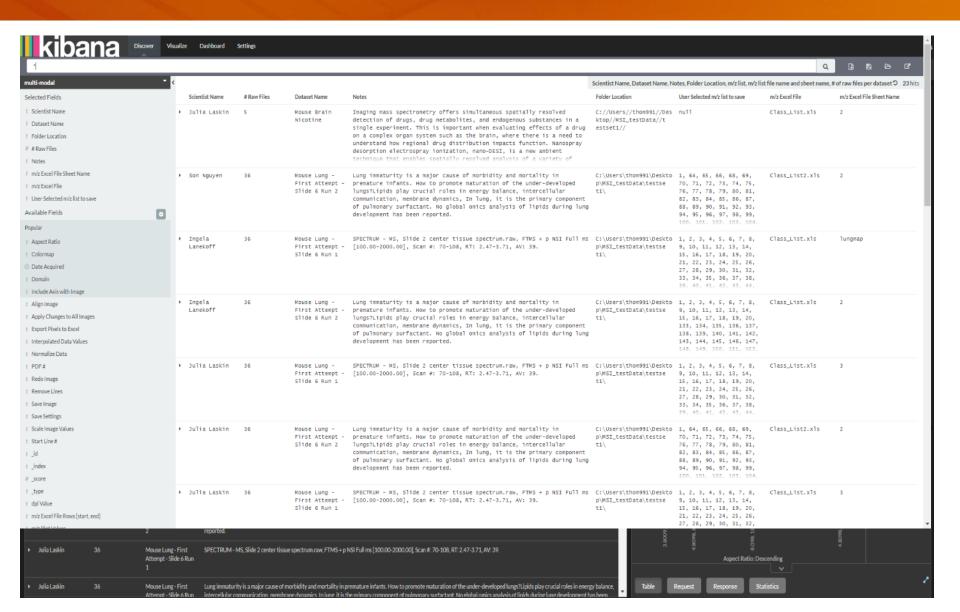


ELK Stack











Proudly Operated by Battelle Since 1965

MSI QuickView Web App

→ C f [] localhost 8080/Java Matlab Integration/ 20 BB B B B B B **MSI QUICKVIEW** Intensity-Time Spectra Intensity-m/z Spectra 0.11 0.0010 0.000) / + 0 - / x Ion Image Panel title / + 0 - / ×



Summary

How is this helping ...

- MSI QuickView
 - Near-real time analysis
 - Visualization and Analysis on normal laptops
- ProvEn
 - Experimental reproducibility
 - Re-executable workflows
- ELK Stack
 - Openly available
 - Breaks down complex information
 - Tailored information representation
 - Querying capabilities

Pacific Northwest NATIONAL LABORATORY Proudly Operated by Battelle Since 1965

Future Work

- Extend MSI QuickView to multi-modal datasets
 - Liquid extraction surface analysis (LESA)
 - Electrochemical Microscopy
 - Flourescence Microscopy
 - Hyperspectral Microscopy
- Port current desktop application capabilities to the web application
- Automate metadata capture workflow for new datasets into the ELK stack
- Investigate the use of ProvEn Platform as a message database

Team



MSI QuickView

Mathew Thomas Nhuy Van

MSI Imaging

Julia Laskin Son Nguyen

Provenance

Todd Elsethagen Bibi Raju Eric Stephan

Acknowledgement

- Chemical Imaging Initiative PNNL LDRD
- LungMAP Grant No. NIH/NHLBI U01HL122703
- NIH R21 NIEHS Grant No. 1R21ES024229-01
- □ Integrated End-to-End Performance Prediction and Diagnosis for Extreme Scientific Workflows (IPPD)